Encouraging users to improve password security and memorability

Article (Published Version)

International Journal of Information Security
https://doi.org/10.1007/s10207-019-00429-y

REGULAR CONTRIBUTION

Encouraging users to improve password security and memorability

M. Yıldırım[1] · **I. Mackie**[2]

Abstract
Security issues in text-based password authentication are rarely caused by technical issues, but rather by the limitations of human memory, and human perceptions together with their consequential responses. This study introduces a new user-friendly guideline approach to password creation, including persuasive messages that motivate and influence users to select more secure and memorable text passwords without overburdening their memory. From a broad understanding of human factors-caused security problems, we offer a reliable solution by encouraging users to create their own formula to compose passwords. A study has been conducted to evaluate the efficiency of the proposed password guidelines. Its results suggest that the password creation methods and persuasive message provided to users convinced them to create cryptographically strong and memorable passwords. Participants were divided into two groups in the study. The participants in the experimental group who were given several password creation methods along with a persuasive message created more secure and memorable passwords than the participants in the control group who were asked to comply with the usual strict password creation rules. The study also suggests that our password creation methods are much more efficient than strict password policy rules. The security and usability evaluation of the proposed password guideline showed that simple improvements such as adding persuasive text to the usual password guidelines consisting of several password restriction rules make significant changes to the strength and memorability of passwords. The proposed password guidelines are a low-cost solution to the problem of improving the security and usability of text-based passwords.

Keywords Text-based password authentication · Password memorability · Password security · Information security · Password policy · Password composition rules

1 Introduction

Authentication is one of the most important areas in computer security, and the use of traditional text-based passwords has been well studied. However, this type of authentication mechanism has drawbacks. Various alternative authentication schemes which aim at aligning security and usability have been proposed. These proposals range from graphical password authentication to location-based authentication [22,26,53]. However, none of these schemes could overcome the simplicity and affordability of typing a sequence of keyboard characters to allow authenticating users [4]. As a result, traditional text-based passwords are still the most popular authentication mechanism on the Web, and they are likely to remain so in the near future [30].

Unfortunately, in terms of usability, text-based password authentication is quite problematic. A good password needs to be "easy to remember and hard to guess" at the same time, as suggested by Wiedenbeck et al. [59]. However, passwords which are easy to remember are generally short or based on dictionary words (or slight variations). Therefore, these passwords become vulnerable to dictionary attacks. Passwords including personal information are also memorable, but they risk to be guessed by people close to the password owner and attackers that have collected information about the user.

Passwords are considered one of the most significant risk factors in terms of security in information systems as they are vulnerable to attacks [8]. This vulnerability is mainly due to user behaviours and practices and not related to the password system itself. The main problem arises from the memorability issue which ultimately causes the other problems related to passwords such as reusing, sharing and choosing weak passwords. These problems are well known, and they are called

✉ M. Yıldırım
merve.yildirim@erzurum.edu.tr

[1] Erzurum Technical University, Ömer Nasuhi Bilmen Mahallesi, Havaalani Yolu Caddesi No:53, 25050 Yakutiye, Erzurum, Turkey

[2] University of Sussex, Brighton, UK

Published online: 11 April 2019

'the human factor problems' by researchers in the password authentication domain [30,52].

Most people are actually aware of the importance of choosing strong passwords to protect their information. However, there is a lack of password creation advice or guidelines to help and motivate them to compose strong passwords. It is commonly stated that a password should include a mix of keyboard characters and should not include meaningful words from dictionaries [62]. Similar advice can also be found in websites, policy papers and many other password security themed articles. Unfortunately, despite these guidelines and advice, most users do not adopt good security behaviours and consequently choose weak passwords. Based on occurrences of security failures and the difficulty to determine which particular rules measure the importance or necessity of strong password creation, research shows that users commonly underestimate the risk associated with weak passwords [19,65].

Shay et al. [50] indicated that users often appear to lack motivation to produce strong passwords as they are not convinced of the importance of suggestions given in the guidelines. It was shown that users' awareness of the problems is not enough to dissuade them from adopting undesirable security practices such as using dictionary words, sharing and reusing their passwords. So, more effective approaches are needed to convince users to behave in a secure manner in the password authentication domain. Rather than only telling them why they should choose strong passwords or restricting their choices with tedious password policy rules, showing how to create better passwords in efficient and fun ways is more convenient.

As described above, human factors play a key role in password security. However, security problems caused by user behaviour have not been totally solved. As previous studies proved that existing password policy rules are not adequate to motivate users to choose strong passwords, this study presents the idea of including several password creation methods in password guidelines and also adding motivating elements to the password creation process without enforcing any restriction rules. Since traditional methods of imposing excessive restrictions have not been very successful, it is suggested that a system that subtly persuades users and offers concrete advice may be more successful. Thus, this study explores whether motivating users with an effective password advice and useful instructions to create strong and memorable passwords is better than obliging users to apply strict password policy rules. This paper recalls the previous studies on password guidelines and reports on an empirical study that has been carried out to evaluate the efficiency of the proposed password guidelines by comparing it with the usual password policy rules.

Overview The rest of this paper is structured as follows. In the next section, we discuss the background and some related work. In Sect. 3, we give the methodology including the details of the empirical study including the design, measurements, apparatus and procedure. In Sect. 4, we give the results together with an analysis. In Sect. 5, the findings are discussed, focusing on the efficiency of proposed password guideline on improving the strength of the users' passwords as well as the compliance behaviour. Finally, we conclude in Sect. 6.

2 Background and related work

To increase the strength of user-chosen passwords, users are typically required to adhere to a set of rules known as password guidelines when creating passwords. Users compose their passwords following the specific requirements given in the guidelines. For example, the password must contain at least eight characters including at least one number or one upper case letter, and it should not contain the username. There are various password guidelines that are used by organisations, and they should be written efficiently to provide adequate security levels in the organisations [7,52].

According to a study, a user has on average 25 online password-required accounts and uses eight passwords per day [19]. However, nowadays, users may even have much more than 25 passwords. As users are expected to use different passwords for each account to avoid security failures, it is difficult for the brain to remember many discrete sets of illogical and random bits of information and then associate each set with which account. The user's response to this situation is generally adopting strategies such as choosing weak passwords or writing them down, which ultimately undermine the security of the systems they use [36]. Some methods are used to replace this subversive behaviour with appropriately suitable behaviour for authentication [60]. These methods aim to direct user behaviour by implementing strict password creation guidelines [32], proactive password checkers [63] or password expiry [66], to ensure a high security level. In addition to these, password management systems are also used nowadays. To save many passwords in a system may be a solution for the memorability problem, but it is also problematic regarding security since the protection of all passwords depends on one single password named "master password". If the master password is cracked, then all user's passwords are obtained. Recent research shows that these advices, measures or system features do not always work as expected. They sometimes have negative effects upon usability and security, contrary to designers' intentions. Where users are given unreasonable constraints, they may more likely adopt insecure workarounds which are easy to use for them [47]. As it is well known, users mostly don't follow the strict security guidelines prescribed within authentication schemes [68]. Both system administrators and end-users struggle to bal-

ance the security and usability of the authentication system. This shows that the current forms of text password schemes which are unable to offer solutions to current socio-technical authentication problems have to be abandoned in the future [39]. Thus, it is inevitable to reform the text-based authentication schemes.

Previous research showed that strict password policy rules do not increase the password security [32,38,52]. As previously mentioned, users adopt coping strategies such as writing passwords down or sharing them when they are forced to use those rules to compose their passwords. However, it might be possible to create strong and memorable passwords by changing the content of password guidelines.

The content of password guidelines should be created carefully as it is important to provide suggestions and instructions to users on how to create good passwords. Grawmeyer and Johnson [28] conducted a study to investigate users' password generation behaviour. All the passwords estimated as highly secure and secure in the study were in fact insecure passwords containing a single word. Therefore, the authors suggested that password guidelines contained in security policies should be devised and founded on a sufficient theoretical understanding of the users' task.

Since only a few studies have been conducted about the construction of password guidelines so far, there is lack of empirical data on the guidelines and passwords which were created complying with them [38]. For example, there is not a sufficient number of experimental studies conducted to evaluate the NIST guidelines [6] which are used to produce password composition policies. They are still mostly based on theoretical estimates. Zakaria [64] conducted a laboratory-controlled experiment to test the compliance to NIST guidelines by different experimental groups which were given different persuasive rationales based on their personality differences. Although it is interesting to investigate the effect of personality variables on susceptibility of users to persuasion, it is difficult to make a definite conclusion. Evaluation of efficiency and security of some other guidelines are also based on very small-scale laboratory studies [47,57].

The updated NIST guidelines regarding authentication systems were released in June of 2017 which include new recommendations for managing and accepting user passwords including password length, complexity, blacklists, rate limiting and two-factor authentication systems [27].

The new guidelines suggest not to impose any password policy rules except the minimum length restriction so as to improve usability and provide a user-friendly authentication system. According to the updated guideline, the password generated by users should be at least 8 characters long; the authentication mechanism should allow users to use long passwords and paraphrases (up to 64 characters) and disallow the passwords in blacklist which were compromised before. There are more innovations such as allowing copy-paste and display the passwords feature, disallowing the password hints and eliminating the password expiration without reason. These are beyond the interest of this research study, as it focuses on eliminating the password composition rules. Since the latest guideline is relatively new, it has not been applied by most organisations yet. Therefore, there is not enough studies conducted to evaluate the efficiency of new recommendations. This study, however, puts the NIST's "no password policy rules" recommendation into practice providing several password creation methods and persuasive messages to users to create their passwords without imposing any password composition rules. It compares the results with the usual password composition rules.

The few recent research studies related to the new NIST guideline focus on blacklist (users' passwords should be checked if they are compromised as they are already listed in previous breach corpuses) [29,37,48] and two-factor authentication [16]. However, as far a we know, no studies have yet been conducted to determine the usefulness of the new guidelines regarding password complexity. As the new guideline is released just a year ago, the suggestions have yet to be implemented by many application developers. This study is the first one which takes into account the new NIST guideline applying the "no password policy rules" recommendation and evaluating its efficiency. Besides, abolishing the use of strict password policy rules may not be enough to encourage users to create strong and memorable passwords. By providing useful methods on password creation stage, they should be directed to do so.

Therefore, the purpose of this study is to guide users to create stronger passwords to improve password security providing several password creation methods in password guidelines instead of the usual password policy rules. People who are particularly interested in doing research to increase the strength of their passwords and create complex passwords can find many password creation methods and tips on the web to help themselves. However, since most users tend to ignore security precautions and they are not interested in doing research about security, including these methods in password creation process, the guidelines should be more influential and efficient. Thus, they would see how to create strong and memorable passwords correctly before they create their own. The methods provided to users in the study inspire users to create their own formula which they can use to turn any simple word to a complex password.

This study also investigates the effect of a persuasive text provided to users on the password creation process to encourage them to create their unique password creation formula on password strength. The existing password guidelines only focus on providing information on how to compose a good password. According to Cialdini [13], people will more likely comply with a request when they are provided a rationale. Accordingly, a study revealed that a password guideline

including a rationale as to why choosing a strong password is important indeed improved the password compliance [64]. However, this study aims at not only telling people what is important and what should be done to increase password security but also show them how to do it. Therefore, users are provided a persuasive message along with example methods. In fact, one of the most important components of persuasion is free choice [46]. Accordingly, the proposed password guideline allows users to be free to create their own strategy on composing passwords. It means that the example methods aim to inspire users to create their unique encryption formula.

In summary, this study seeks to discover if users will create strong and memorable passwords when they are provided several password creation methods along with a persuasive message, without imposing any password policy rules.

2.1 Password creation policies

Password restriction policies are a series of rules which determine the content and format of the passwords accepted by an authentication system. These policies are used by system administrators to enhance computer security by guiding users to create more secure passwords.

In 2006, the National Institute of Standards and Technology (NIST) updated the "Electronic Authentication Guideline" [6] to be used by security system administrators for the implementation of electronic authentication. This guideline provides heuristics to measure the strength and efficiency of a password restriction policy considering bits of entropy to determine a password value's uncertainty. In this guideline, estimation of Shannon's Entropy [49] was used for the entropy calculation. However, several studies [10,14,38,58] have found that passwords created with particular password policies were more difficult to guess than the ones created with the NIST model suggestions.

Komanduri et al. [38] conducted a large web study to compare four different password restriction policies. They found that users have less difficulty to comply with creating a 16-character minimum password compared to an 8-character minimum excluding dictionary words or further restrictions. In addition, passwords with at least 16-characters provide the best security. They also measured the password strength using a calculation entropy [50], and thus, they showed some misconceptions about how restriction policies affect password strength. Their findings conclude that adding digits significantly increased the entropy of passwords, but excluding dictionary words increased the entropy less than expected. Also, the findings showed that passwords created by users barely exceeded the minimum requirements.

Contrary to what is believed, some researchers have claimed that password restriction policies do not improve password security [1]. There has been some laboratory [57] and field studies [34] conducted to test this claim. Results show that it is difficult to create and remember passwords for users when they are enforced to employ strict and complex password policies. To cope with remembering difficult passwords, users commonly adopt insecure password practices. These policies also help attackers to guess passwords more efficiently as they can decrease the number of candidate passwords based on the restriction policy. In their website study, Florencio and Herley [20] found that users only tolerate the restriction policies if they have no other choice. However, most of the systems use policies requiring passwords 20-bits strong. This causes a burden on users to deal with a cumbersome password restriction policy. The authors also note that websites which typically users do not care too much to create strong passwords to log in are often the most popular ones and also the most likely to be attacked as they have great amounts of assets for hackers. If such popular websites continue to force strict password policies, it might increase the security slightly in exchange for a considerable usability cost.

Finally, NIST guidelines have been updated in 2017 emphasising the uselessness of strict password policy rules on creating strong passwords [27]. Therefore, it recommended organizations not to require users to apply these rules anymore. It stated that the length of the password affects the strength of the password the most, so it suggested the use of long passwords and passphrases. Accordingly, this study introduces a guideline including several password creation methods and a persuasive message without imposing users any password policy rules to compose their passwords. The new guideline aims to motivate users to create cryptographically strong and also memorable passwords.

2.2 Password creation advice

Most systems that impose password restrictions offer their users password advice about creating passwords. The purpose of password advice makes adoption of password policy rules easier and also motivate users to create stronger passwords. In a study, password practices of ten popular Internet sites which enforce password policy rules and offer password advice were examined [23]. That the websites' password restriction policies and password advice are vastly different, sometimes caused conflict between them. In most of the websites, password advice was found ambiguous and unhelpful by users. As existing password policies and advice are far from being consistent and effective, it is not easy for users to form accurate mental models of how to create a secure and memorable password.

Murray and Malone [43] recently highlighted the characteristics of the password advice distributed by different organizations. They found out that there are substantial discrepancies between advice used in different environments. Websites enforce different password creation restrictions.

Their research also showed that some advice stated as the best practice by security researchers is even not included in the majority of advice. This contradiction may cause users unwillingness to follow advice.

2.2.1 Mnemonic passwords

There is a wealth of research investigating the best way to advise users to create secure and memorable passwords. In an attempt to encourage users to create easy-to-remember passwords, mnemonic phrase-based passwords have been first proposed by Barton and Barton [2]. Mnemonic passwords are derived from a memorable sentence where users generally use a letter of each word in the sentence. Although most of the password advice research is about mnemonic passwords, they are rarely recommended to use in practice [23].

There are more studies which present the different ways to generate a mnemonic password. Vu et al. [57] used two mnemonic password generation methods in a user study, and let all users choose their own sentence. As the passwords created with the mnemonic string method typically have more characters, they were thought more secure. However, the authors found little difference in password creation times, login times and recall error rates between two methods. In a previous study, they had also found that passwords which contain more characters were more resistant to cracking [47]. Unfortunately, as the way of substitute words and characters suggested in the study is well known by attackers, mnemonic string method may not be very much secure as previously thought.

2.2.2 Password chunking

Very little research has been done in password advice apart from mnemonic passwords. There are studies on the use of chunking [17,41] to help users to create easy-to-remember passwords. Carstens et al. [9] performed a field study applying chunking theory to an organisational password guideline. They compared common password advice (that the password should contain at least 7 characters, be a combination of symbols and letters, not contain repeated characters more than twice, not be a dictionary word or personal data), to two-chunk, three-chunk and four-chunk passwords. The authors found that four-chunk passwords were not only longer, but also more memorable than the 7-character, two-chunk or three-chunk passwords. However, they did not carry out a security analysis of the created passwords, or the password advice itself. Therefore, four-chunk passwords may even have been less secure, as they contained fewer distinct symbols than other passwords. Furthermore, since participants were explicitly told what to use for their chunks such as participants' first and last initials, spouse's initials, employment start date in the password guideline, it would not have been difficult to predict participants' passwords for an attacker, particularly the one familiar with the user.

2.2.3 Password strength meters

Password advice can also be represented with a tool measuring strength of the password and giving users a numerical result or statements such as 'weak', 'strong' and 'very strong'. These tools are called "password strength meters" which typically illustrate the strength of the currently chosen password when a user is registering for an account. The meters are commonly used by popular websites (Gmail, PayPal and eBay). In an online user study conducted with over 2000 participants, different password strength meters were evaluated [55]. The results showed that the passwords created by users who used password meter were more difficult to guess than the passwords created by users who did not use a strength meter. Furthermore, users created much stronger passwords when they used stringent password meters. However, the authors found that meters which are too stringent may cause users to lose motivation and ignore the meter. In another study, a novel method called adaptive password strength meters (APSMs) were proposed to measure password strength [10]. Adaptive password strength meters use Markov models [40] to measure a password's strength as the collective probability of each character following the previous characters in the password. These probabilities can be calculated based on either a training set of passwords or the passwords currently in use. Although the authors claim that APSMs are better than any other proposed password strength metric to date as it can score passwords closer to the "ideal" password strength meter, there has not been conducted any formal usability study of APSMs and a practical security evaluation. Kelley et al. [35] introduced different calculators for estimating the number of guesses required to crack a password using a particular cracking algorithm. They calculated the percentage of passwords which can be cracked with the implemented algorithm given a number of guesses. To compare cracking performance across algorithms, guess number calculators for several cracking algorithms on the same set of passwords can be implemented. Guess number calculators may be considered the more practical and efficient method of proactive password checking [3] than running a computationally intensive password cracking algorithm [22].

2.3 Usability of password security

Passwords authentication mechanisms mostly involve a trade-off between security and usability. The main reason of the imbalance between them is memorability problem. When users are forced to create long, complex and randomly generated passwords, they are likely to write them down or forget

them. On the other hand, when users choose the weak and predictable passwords, they are susceptible to attacks. The researchers who tried to crack passwords conducted several experiments and the results proved the weaknesses of the user-chosen passwords [25,36]. It seems more secure password means the less usable password or vice versa [5]. Some authors investigated the relationship between password security and usability by conducting several studies [31].

In relation to this study, the next section discusses the memorability criteria of usability causing the security-usability contradiction.

2.3.1 Memorability

Memorability is the most important issue in knowledge-based authentication systems considering the limitation of human memory that puts systems security into high risk. Many studies pointed out the users' difficulty in remembering passwords [1]. Users typically adopt coping strategies to avoid forgetting and resetting passwords.

Vu et al. [57] tested the memorability of text passwords which are created obeying various password policy rules. They found that remembering five passwords is more difficult than remembering three passwords. Also, users tend to create passwords which are obviously connected to the accounts, as a memory assistance coping strategy.

Chiasson et al. [12] conducted a study to compare the memorability of multiple text passwords and multiple PassPoints graphical passwords (a PassPoints password is a sequence of points, chosen by a user, on an image). They found that after the passwords were created, graphical passwords were much more easily remembered than text passwords. As remembering different passwords across accounts is challenging for users, they commonly use coping strategies to overcome the memorability issue. One of these strategies is choosing similar passwords across accounts which causes multiple password interference. This issue has been studied in a few other graphical passwords-related research papers [11,18,42].

2.3.2 System-assigned and user-chosen passwords

In text password authentication mechanisms, either users are allowed to set their own passwords or they are assigned a typically random password by the system itself. Generally, system-generated passwords are much more secure than user-chosen passwords, since users mostly choose weak or predictable passwords to remember easily. Also, most users are not aware of the probability of guessing attacks and capabilities of attackers to compromise passwords [1]. However, passwords assigned by the system are harder for users to memorise and remember as they generally consist of combinations of unrelated characters [62]. Therefore, users would not have the chance to use cognitive methods which help memorability in creating passwords. Also, passwords that have no meaning for users consequently make them harder to remember [57,68]. A research study proved that memorability of system-assigned passwords and system-assigned passphrases is equivalent regarding password strength [51]. This means that remembering system-assigned passwords is a great burden on human memory without any memory aid provided by the authentication system. There are other methods which have been researched to aid text password memory such as using semantic content [33,54] and cueing. However, Wright et al. [61] found that recognition passphrases which are entered by users by selecting the assigned words from a list are not more memorable than system-assigned passwords. This proves that some forms of memory assistance are not sufficient, so further studies are needed to evaluate effective forms of memory aid.

In the next section, we describe the empirical study which was conducted to evaluate the proposed idea.

3 Methodology

3.1 Introduction

A web-based empirical study was carried out to investigate the assumption that users can create stronger and more memorable passwords if they are not enforced to comply with strict password policy rules. The study evaluates the effect of several password creation methods on the strength and memorability of users' passwords. It also explores the users' password behaviour and practices in either cases: with and without following strict password policy rules.

The proposed password guideline was established based on the idea of allowing users to create their own password creation formula would increase the security and memorability of the passwords. This is because following predetermined strict password policies causes users to adopt coping strategies as suggested by previous works. In the proposed guideline, three methods were provided to the users as an example. While determining these sample methods, first previous password guidelines and advice were scrutinised and then applicability, adoptability and usability of the methods were considered. With these methods, it is aimed to give users an idea to compose their passwords. Users either would adopt them as they are and produce passwords using their self-chosen words or modify the methods slightly. They would also be inspired by the methods and create their unique password creation formula which is the main purpose of the proposed guideline. In addition to the sample methods, a persuasive text is also included in the proposed guideline which is considered an effective way to motivate users to create strong passwords.

Ethical approval was obtained from the University of Sussex to perform this study. The following section presents the details of this study.

3.2 Methodology of the empirical study

3.2.1 The design and apparatus

The empirical study used the between-subject design where participants can be part of the experimental group or the control group, but cannot be part of both. There is one control group and one experimental group in the study. The strength and memorability of the passwords created by participants in each group were measured and compared to each other. While the participants in the control group were given some password policy rules to be followed when creating their passwords, the participants in the experimental group were given several password creation methods as examples to create their own formula to compose passwords without having to follow any rules.

Two different websites were created to collect data for this empirical study: one for the experimental group and one for the control group. The apparatus used in this study is as below:

- A password guideline including five password composition rules for the control group.
- A password guideline including three sample password creation methods and a persuasive message and important notes for the experimental group.
- A password register/login page of a web site for the control group.
- A password register/login page of a website for the experimental group.
- Two sets of questionnaires: one for the control group and one for the experimental group.
- Consent forms to read and accept for all participants.

Once the participants in each group had signed up, they were provided with a message and a survey link directing them to another website to fill out a questionnaire. While the questionnaire given to the control group included 19 questions, the experimental group's questionnaire included 30 questions. The questions in the control group's questionnaire were common for both group. The average time to complete the questionnaire was approximately 15 minutes for the experimental group and 10 minutes for the control group.

As stated above, the password guidelines which were distributed to the participants are different according to the group to which they were assigned. The participants in the control group received a set of password composition rules are as below:

The password composition rules for the control group:
- Your password must be at least 8 characters long
- Your password must contain at least one upper case, one lower case, one number and one special keyboard character
- Your password should not contain your username
- You should use different passwords across different accounts
- Your password should not be easily guessable

The participants in the experimental group were given a persuasive message telling them that it is possible to create strong and memorable passwords applying some methods. The message attempts to persuade participants to create a unique formula; thus, they could turn even a simple word to a complex password which is hard to crack. In addition to this message, participants in the experimental group were also given these methods with examples. The password guideline of the experimental group was framed using logical reasoning by providing explanations such as the fact that if users create weak passwords for ordinary websites, a crafty hacker can obtain that password easily. If using the same or similar password is the user's habit, it would not be difficult for the hacker to guess the other passwords created for important accounts such as bank account.

The password guideline including the persuasive message and password composition methods given to the experimental group are as follows:

The participants were shown a register/login page to enter their username and password. At the beginning of the empirical study, the participants were assigned a unique ID number. The final apparatus involved in the study is the questionnaires of the control and experimental group. The control group's questionnaire contains several questions on demographic details of the participants and some questions related to password constructions and usage. In addition to these questions, experimental group's questionnaire contains some questions to find out the user satisfaction with the given methods and the effect of the methods on users' password choice.

The questions in the questionnaires were set to understand the reasons for the participants' password practices and determine the influence of proposed password guideline on password strength and memorability. A pilot survey was conducted before with fewer participants to test the quality of the questions. The content of several questions in the preliminary questionnaires was changed to better suit the aim of the study. Also, some previous studies were reviewed that used the several materials and questions [50,64]. Not only the survey questions but also the proposed password guideline including three password creation methods were tested with a small number of participants before the main experiment. In this way, the methods were examined whether they are readable and applicable easily.

You can create very strong passwords with simple and memorable methods to protect your accounts! Please read carefully and understand the logic or bases of the methods
Before signing in, please use the methods given below to create your password. These methods offer a high level of security by leading you to create your own encryption formula to produce strong passwords from memorable words. Once you have understood the encrypting logic behind the manipulation of memorable words with equally memorable strings of numbers to create your very own formula, you will be able to transform words and numbers that are memorable for you, into very strong passwords. Memorable words mixed appropriately with memorable strings of numbers leads to strong passwords.
Important Note
You should choose words and numbers unrelated to your personal details. For example, don't encrypt your name or surname. And don't use strings of numbers that indicate your date of birth. Even though others do not have a clue about your formula, it is risky.
Remember that password selection is very important to protect your confidential data. Without exception, you should care about your password selection for anything from ordinary websites to bank accounts. A crafty hacker can easily obtain your weak password within seconds. Also remember that if you use very similar passwords across different accounts, and, once the hackers have obtained one of your passwords, they can easily guess the others.
Below are examples of workable methods. You should not apply exactly any of the examples as your own password. The examples are given only to show the possibilities of good encryption formulation that consequently lead you to think of your own encryption formulation resulting in a memorable formula.
Method-1
Step 1: Pick a word. Let's say, "education" as our plain password.
Step 2: Specify a number. Let's say, 347. So we have the word "education" and the number "347". Let's encrypt them.
Step 3: Convert the 3rd, 4th and 7th letters of the word "education" to upper case. We now have "edUCatIon".
Step 4: Place the numbers 3, 4 and 7 after each of the upper case letters. This gives us "edU3C4atI7on".
Step 5: Change the value of each of the numbers in Step 4 by increasing or decreasing each. In this case, we will choose to increase each by 2. Therefore 3 + 2 becomes 5, 4 + 2 becomes 6 and, 7 + 2 becomes 9. So now we have the strong password "edU56atI9on".
That's it! It is almost impossible to guess and very hard to crack. You can even write the plain password somewhere to help you remember it. As long as no one knows your formula that converts a plain password into a strong password, plain passwords are meaningless to them. Here's another example. Let's pick a Turkish word and the number 148. Our plain password here is "bilgisayar." When we applied the same formula, this plain password converts into the strong password "B3ilG6isaY10ar".
Method-2
Step 1: Choose a string of plain numbers. Let's choose the numbers "12345".
Step 2: Specify a combination of letters and keyboard characters. Let's specify "`m_y_`". (Letters separated by underscores).
Step 3: Mix the string of plain numbers with the combination of letters and keyboard characters. In this case, we sequentially alternate the individual numbers of Step1 with the letters and keyboard characters of Step 2. Thus, we get the strong password "`1m2_3y4_5`".
Method-3
If you want to use meaningful words and phrases you have to create a very long password combining letters, numbers and other keyboard characters. For example, "`myfavouritechicredshoes-size4`." This phrase is meaningful to you, so you can remember it easily, but for other people it should be hard to guess. You can combine unrelated words which you can associate. An example of this is "`elephant.zoo.travel.Africa`". An elephant might remind you of a zoo and a travel to Africa.
Important Notes
– You can pick a related simple password or add some more characters to your encrypted password to remind you of the site for which you create the password.
– You should not use the same examples and/or formulas given in the above methods. You should create your own.
– You should use different passwords for different accounts.
– You should never share your passwords and/or your formula with anyone.
– You can apply your formula to different words and numbers to create different passwords.

3.2.2 The procedure

The participants were informed about the study and recruited using social media posts and flayers hanged on school boards in several universities in UK, USA and Turkey. Links of both websites were posted on social media platforms, and potential participants were asked to choose only one of them to sign up. The place of the links was changed on different posts and flayers as users might tend to click or type the first one. In the last days of the experiment, only the link of the website which has lesser signed-up participants was shared to recruit more people. This helped to avoid the imbalance between the numbers of participants in the control and experimental groups. Also, only university students' responses included in the empirical study.

At the beginning of the empirical study, all participants were automatically assigned a unique ID number. These ID numbers were used to match the participants' credentials and questionnaire responses. The participants in both groups were given a brief information about the study and asked to read and accept the consent form to participate. Once they had accepted the consent form, they were able to sign up their website. For those participants who were interested in getting more information about the study researcher's, contact

information was provided in the consent form. As presented above, participants received different password guidelines while signing up. After the participants signed up the websites successfully, they were asked to click the provided link which would direct them to the survey website to fill the questionnaire. Before filling the questionnaire, a brief introduction was provided to participants informing them about the approximate time which the questionnaire would take. Finally, the empirical study was finished with a message of thanks to participants for their participation.

The participants were asked to login again after a week and a month to find out whether they recall their passwords. The procedure of the study was exactly the same for both groups except the password guidelines and additional questions in the experimental group's questionnaire.

3.2.3 The measurements

There are several measurements involved in the empirical study: password strength, password length, memorability, password policy compliance, use of the given methods, user satisfaction and persuasiveness. The password strength was measured using tools known as the "Password Meter" [45] and "How Secure is my Password?" [15].

"Password Meter" measures the password strength using a combination of several important attributes that constitute a particular password, such as length (i.e. number of the characters), the frequency of uppercase, lowercase, numerical characters and alphanumeric characters (i.e. punctuation and mathematical symbols). Password strength calculation is made by adding points if the password meets the requirements, and deducting points if not. The tool categorises the passwords by giving scores where the maximum score is 100% according to complexity of the password. The passwords will be categorised as follows: very weak ($0\% \leq$ password score $< 20\%$), weak ($20\% \leq$ password score $< 40\%$), good ($40\% \leq$ password score $< 60\%$), strong ($60\% \leq$ password score $< 80\%$) and very strong ($80\% \leq$ password score ≤ 100).

"How Secure is my Password?" measures the time in which the password entered could possibly be cracked by a computer. The tool takes into consideration the length, whether the password given looks like a dictionary word and the character variety while measuring the estimated cracking time.

Using both tools, participants' passwords were measured. Based on results of the measurement, passwords were evaluated according to their strength level and possible cracking time.

The next element that was also measured in this study is the memorability. To evaluate the recall rate, the participants were asked to login to the websites after a week and after a month. To carry out the measurement of memorability, participants were given numbers where 0 indicates that user could not login successfully, 1 indicates that user logged in successfully in the first attempt and 2 indicates that user logged in successfully after a few attempts.

The other measurement in the study was password policy compliance. Since the participants in the control group must follow the given password policy rules, the measurement was conducted to find out whether the participants in the experimental group applied these rules to their passwords. Passwords were given scores for each requirement where 0 indicates that the related rules had not been applied and 1 indicates that the related rules had been applied by the participant. Briefly, the measurement showed how close each participant in the experimental group followed the requirements given in the password guideline of the control group.

Use of given methods, user satisfaction and persuasiveness were measured based on questionnaire responses of the participants in the experimental group to evaluate the password guidelines given them.

3.2.4 Demographics

308 people participated in this study. 152 of them were in the experimental group, and they created passwords after reading a persuasive text on how to create strong and memorable passwords. 156 people were in the control group where they had to follow five commonly used instructions to create a password. 142 people in the experimental group and 95 people in the control group filled out a follow-up questionnaire including questions about their password creating habits and their thoughts of given methods. Therefore, demographics and survey analyses were run for 237 participants whereas password strength, compliance and memorability scores were analysed for all 308 participants.

The participants are recruited from college students studying in some universities in UK, USA and Turkey. Undergraduate students as well as the postgraduate students participated in the study. Some of the participants have computer science related background.

There were 111 females and 126 male participants. 90.8% of the experimental group and 86.3% of the control group were aged between 18–35 years. Figures 1, 2 and 3 illustrate the demographic details of the participants involved in the empirical study.

The number of participants who are currently in an undergraduate programme ($n = 108$) and in a postgraduate programme ($n = 129$) was close. Additionally, most of the participants did not have a background in computer science, and only 19% of all participants were from a computer science related major.

This section described the design, apparatus, procedure and measurements of the empirical study and presented the

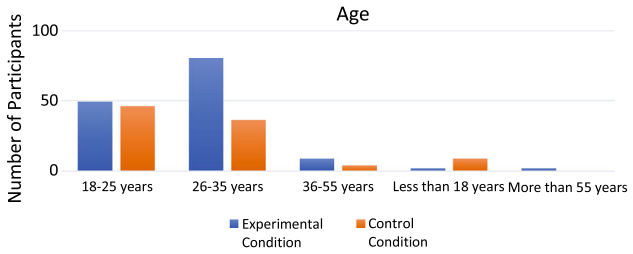

Fig. 1 Age profile of the participants

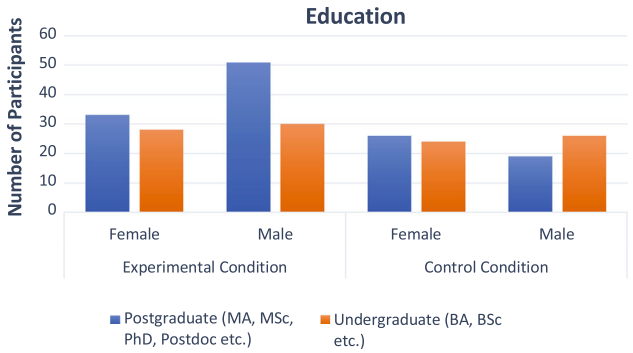

Fig. 2 Education level profile of the participants

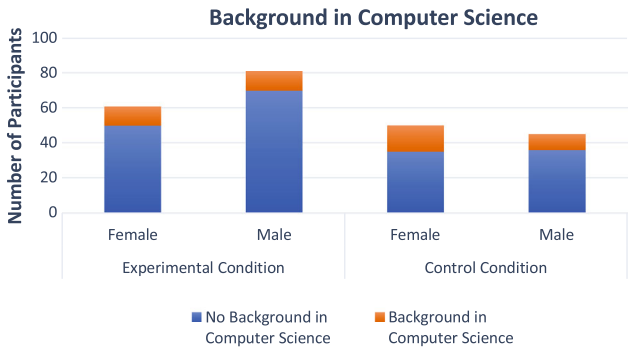

Fig. 3 Education background profile of the participants

full participant information. The next section presents the results and analysis of the empirical study.

4 Results

4.1 The results and analysis of the empirical study

In the following sections, results and analysis of the empirical study are presented in detail.

4.1.1 The Password analysis

Experimental and control conditions were compared in terms of password strength, password length and compliance to the common password policy rules, which were compulsory for control condition. The aim was to see whether people in experimental condition were also following those guidelines unintentionally while creating passwords in line with other methods provided. Table 1 presents the password analysis of the empirical study.

Password strength In this study, password strength was scored out of 100 for each password individually. Though almost the same number of participants in experimental and control groups think they have strong passwords (62.7% vs. 61.1% respectively), an independent samples t test (t (306) = 15.617, $p < 0.001$, $\eta^2 = 0.44$) proved that there was a difference in password strength between conditions (see Fig. 4). The experimental group ($M = 89.08$, SD = 8.84) created stronger passwords than the control group ($M = 71.95$, SD = 10.36). According to the Password Meter, the passwords of the control group and the experimental group are categorised, respectively, as strong and very strong.

According to the results, people are not good at predicting the strength of their own passwords. As Fig. 5 indicates that

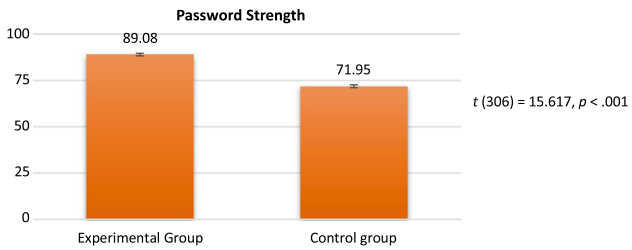

Fig. 4 The password strength of the experimental and control group

Table 1 The password characteristics

Group	Password strength	Password length	Password compliance
Experimental group	$M = 89.08$	$M = 13.39$	$M = 4.74$
	(SD = 8.84)	(SD=4.35)	(SD = 0.91)
Control Group	$M = 71.95$	$M = 9.59$	$M = 6$*
	(SD = 10.36)	(SD = 1.80)	(SD = 0)

*Password compliance score was constant for control group, because it was compulsory
*Password Strength scores were between 0 and 100, calculated with the Password Meter tool; Password Length is the number of the characters used, Password Compliance scores calculated as the total number of rules complied while creating a password, out of 6 rules in total

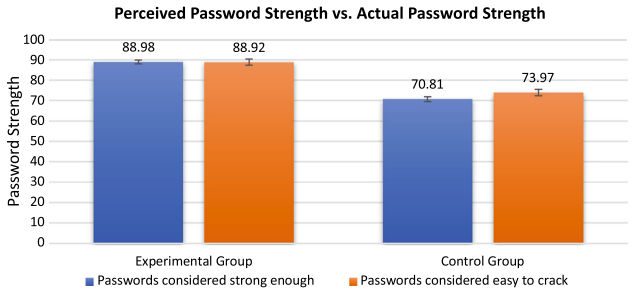

Fig. 5 Differences between users' perception of password strength and the actual password strength

neither in the control group ($t(93) = -1.375$, $p = 0.172$) nor in the experimental group ($t(140) = 0.034$, $p = 0.973$) there were difference in terms of password strength between those who thought their passwords were hard to crack and those who thought their passwords were easy to crack.

Password length Password length was also compared between groups with an independent samples t test analysis. As illustrated in Fig. 6, the results showed that there existed a homogeneity of variance problem, such that Levene's test for equality of variances was significant ($p < 0.001$). However, the t test without equal variances assumption was also significant ($t(200) = 9.986$, $p < 0.001$, $\eta^2 = 0.33$). Experimental group ($M = 13.39$, SD $= 4.35$) created significantly longer passwords than control group ($M = 9.59$, SD $= 1.80$).

Additionally, password length is correlated with password strength. A strong positive Pearson's correlation exists between them, such that the longer the passwords, the stronger they are ($r = 0.775$, $n = 301$, $p < 0.001$). Figure 7 demonstrates the relationship of number of the characters in the password and the password strength.

To check for the interaction of experimental group and password length, password length scores were recoded as a categorical variable with median split method, where the length scores of both groups were divided into two groups: shorter passwords and longer passwords. Then, a 2 (experimental group) $\times$ 2 (password length) factorial ANOVA was computed. This ANOVA yielded significant results both for the password length ($F(1, 304) = 166.079$, $p < 0.001$) and for the experimental condition ($F(1, 304) = 334.695$, $p < 0.001$). However, there was not an interaction of the two variables ($F(1, 304) = 0.649$, $p = 0.421$). Experimental group was better at creating strong passwords both for longer passwords ($M = 95.71$, SD $= 8.84$ vs. $M = 78.56$, SD $= 8.51$) and for shorter passwords ($M = 83.41$, SD $= 7.61$ vs. $M = 67.71$, SD $= 10.36$). Figure 8 compares the strength of short and long passwords of experimental and control group. In this analysis, there was a homogeneity of variance problem as well. However, when the analysis is repeated after data reduction based on outlier analysis with Cook's distance values, neither the results changed, nor the homogeneity problem was solved. Since ANOVA is considered a robust test against the equal variances assumption and since the significance levels reached in the analysis were acceptable even at conservative perspective, these results were considered acceptable. Additionally, since the unequal variances are almost inevitable in this empirical design due to the difference in number of strategies given to the participants in both conditions, no data were excluded from the analyses.

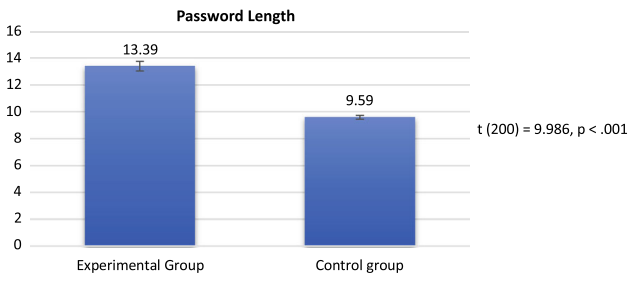

Fig. 6 The password lengths of the experimental and control group

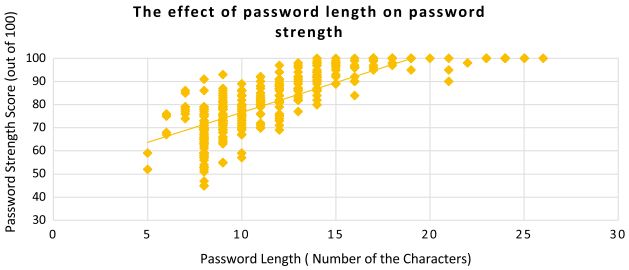

Fig. 7 The correlation of password strength and password length

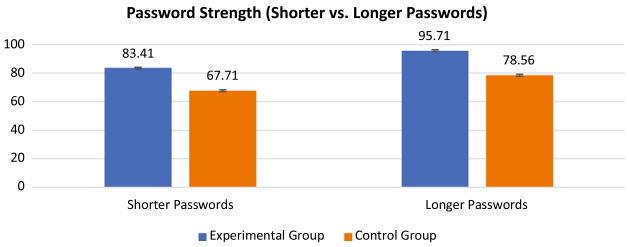

Fig. 8 The strength of the shorter and longer passwords of both group

Password compliance At the beginning of the empirical study, the control group was required to follow a password guideline including the following password composition rules:

The password should contain:

1. at least 8 characters
2. at least one upper case
3. at least one lower case
4. at least one numerical character
5. at least one special keyboard character.

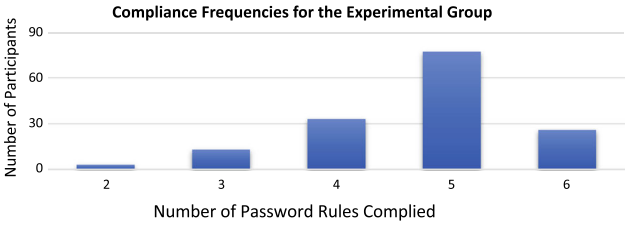

Fig. 9 Password policy compliance frequencies for the experimental group

The password should not contain:

1. The username

Most of the web applications require users to apply at least 4 of these rules while creating a password. However, the participants in the experimental group were not imposed to use them to compose their passwords. Instead, their password guidelines included persuasive elements aimed at motivating users to create passwords which automatically provided these requirements.

To see whether the experimental group also complied with the rules required for the control group, each password was coded in a binary format (1 for compliance and 0 for incompliance). Later, passwords were given a score as the total number of compliance across six items. Although the experimental group's compliance score was significantly different ($t(151) = -17.199$, $p < 0.001$) from 6, the score of the control group, still 89.5% of the participants in the experimental group complied with four or more password creation rules (see Fig. 9).

The experimental group's passwords were stronger than the control group's passwords, which were obliged to comply all the strategies and therefore have the compliance score of 6. This result suggested that compliance to the commonly used password criteria was not a guarantee of strong passwords (see Fig. 10).

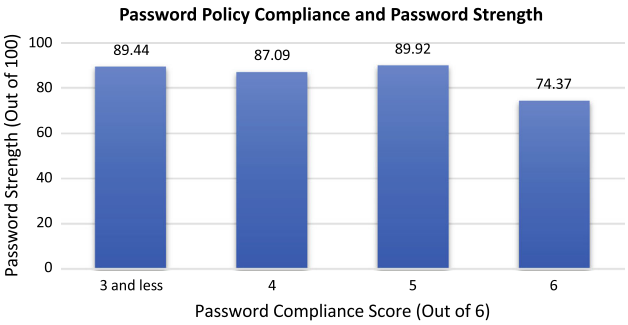

Fig. 10 The effect of password policy compliance score on password strength

The relationship between compliance and password strength was rather complicated. A univariate ANOVA was run to compare password strength, and the compliance scores 2 and 3 were collapsed due to the number of data in each group. So, there was a significant difference between groups $F(3, 304) = 47.905$, $p < 0.001$. A Tukey's post hoc test showed that this difference stems from the highest compliance group, which differed from all other groups with 95% confidence intervals, as shown in Table 2.

Memorability How likely people were to remember a password they create is another central point. In this study, first, memorability of a password after a week and after a month was compared across groups with a 2 (time: week vs. month memorability) $\times$ 2 (experimental group) factorial ANOVA. As shown in Fig. 11, results indicated that memorability significantly decreases as time passes ($F(1, 306) = 59.712$, $p < 0.001$). Memorability after a week ($M = 1.31$, SD $= 0.83$) was significantly higher than memorability after a month ($M = 0.85$, SD $= 0.81$), where a score of 2 indicates correctly remembering a password at first trial, score of 1 indicates correct remembering at many trials, and a score of 0 indicates failing to remember. Additionally, the chances of correct retrieval were higher for experimental group ($M = 1.27$, SD $= 0.05$) than the control group ($M = 0.90$, SD $= 0.05$; $F(1, 306) = 28.320$, $p < 0.001$). This result is in line with experimental group's predictions about memorability, where 86% of participants thought they would remember the password correctly after a week, and 76% of them thought they would remember correctly after a month (see Fig. 12). Hence, no interaction between the experimental group and time was found ($F(1.306) = 0.425$, $p = 0.515$).

To investigate how the conditions were distributed across retrieval levels, two Chi-square analyses were run. For the memorability of a week, the Chi-square test was significant, $\chi^2(1, N = 308) = 15.487$, $p < 0.001$, which meant that experimental group and control group differed from each other at retrieval performances (see Fig. 13). Grouped comparisons supported that the significance stemmed from differences at all levels. In a week period, the experimental group was better at both correct retrieval at first trial (101 vs. 69 participants) whereas the control group either failed to remember more than the experimental group (28 vs. 45) or retrieved correctly at many trials with higher frequency (23 vs. 42).

The second Chi-square analysis showed that the experimental group (57 vs. 25 participants) again performed better at correct retrieval at first trial, $\chi^2(1, N = 308) = 20.147$, $p < 0.001$. Similarly, the control group failed more to remember their passwords correctly (49 vs. 80 participants) after a month.

Use of password creation methods in the experimental group Among the 142 participants who took the follow-up ques-

Table 2 Comparison analysis of the password policy compliance and password strength across compliance scores

Multiple comparisons						
Dependent variable: password strength						
Tukey HSD						
(*I*) compliance	(*J*) compliance	Mean difference (*I* – *J*)	Std. error	Sig.	95% Confidence interval	
					Lower bound	Upper bound
3 or less	4	2.35	3.253	0.889	– 6.06	10.75
	5	– 0.48	2.934	0.998	– 8.06	7.09
	6	15.07*	2.784	0.000	7.88	22.26
4	3	– 2.35	3.253	0.889	– 10.75	6.06
	5	– 2.83	2.222	0.580	– 8.57	2.91
	6	12.72*	2.020	0.000	7.50	17.94
5	3	0.48	2.934	0.998	– 7.09	8.06
	4	2.83	2.222	0.580	– 2.91	8.57
	6	15.55*	1.452	0.000	11.80	19.30
6	3	– 15.07*	2.784	0.000	– 22.26	– 7.88
	4	– 12.72*	2.020	0.000	– 17.94	– 7.50
	5	– 15.55*	1.452	0.000	– 19.30	– 11.80

Based on observed means
The error term is mean square (error) = 114,028
* The mean difference is significant at the 0.05 level

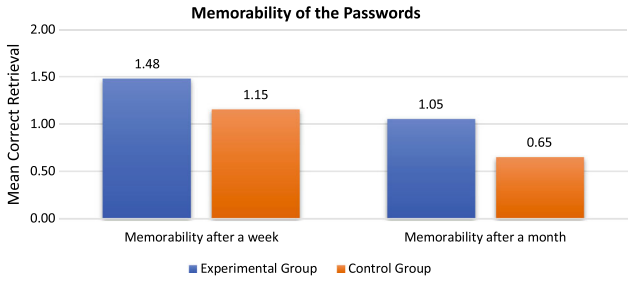

Fig. 11 Memorability of the passwords of experimental and control group

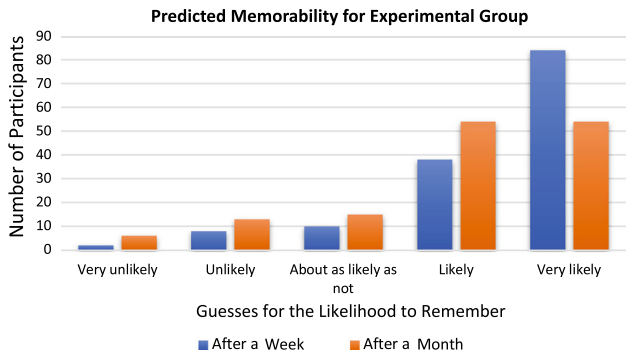

Fig. 12 Users' predictions of memorability in the experimental group

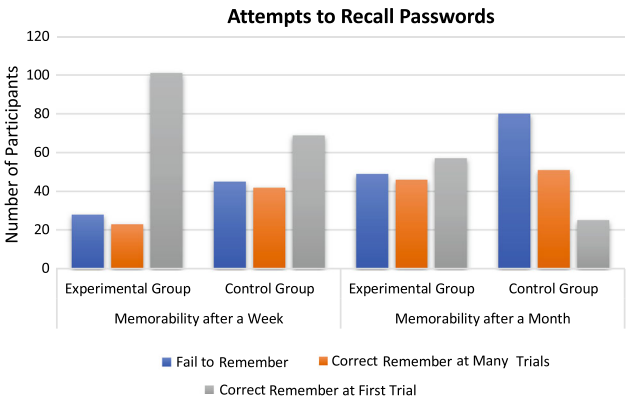

Fig. 13 Attempts to recall passwords after a week and after a month

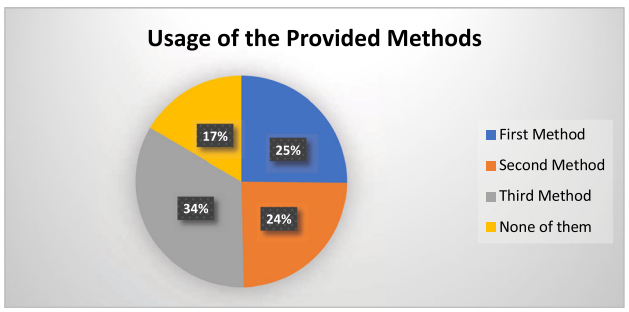

Fig. 14 Preferences for the given methods in percentages

tionnaire, 3 participants were excluded from the following analyses since they selected more than one option at the same time. Among the remaining participants in the experimental group, most of them preferred the third method (34%) (see Sect. 3.2.1 for the details of the given methods). The distribution can be seen in Fig. 14.

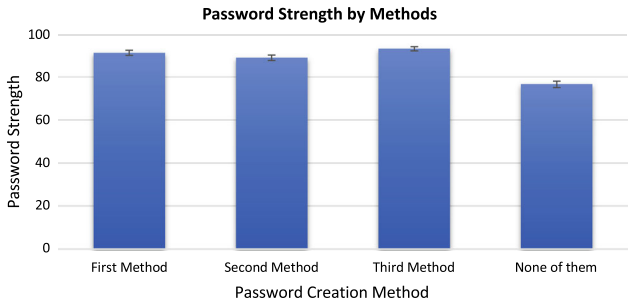

Fig. 15 Password strength by methods

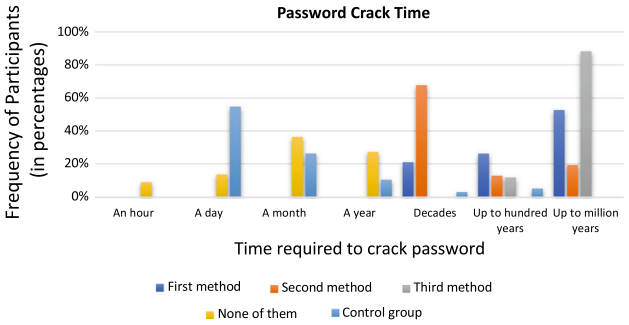

Fig. 16 Estimated password cracking times

Moreover, password creation methods were compared in terms of how strong passwords the participants created by applying these methods (see Fig. 15). A univariate ANOVA showed that passwords created by different methods were significantly different, $F(3, 135) = 30.097$, $p < 0.001$.

To understand which methods are creating the difference, a Tukey's HSD post hoc test was run, and it resulted that the passwords created by applying any of the methods are stronger than the passwords created without applying the given methods (stated as none of them in the figures), with 95% confidence interval (see Table 3). Additionally, passwords created with method three were slightly stronger than the passwords created with method two ($p = 0.047$).

Password cracking times To measure the strength of the passwords in the experimental and control group in another way, a tool, namely "How Secure is My Password," was also used. This tool measures the estimated cracking time of the created passwords. When the passwords were analysed based on the time required to crack them, data indicated that only control group participants and the participants in the experimental group who did not utilise any of the given methods created passwords which are easier to crack (i.e. passwords which could be broken in less than a year). Additionally, when the experimental methods were compared to each other, most of the participants who used method two created passwords to be cracked in decades, as opposed to the participants who preferred the first or the third method. This result is in line with password strength results, which indicated method two created slightly less strong passwords than methods one and three. Estimated password cracking times for both groups' passwords are shown in Fig. 16.

Table 3 Comparison analysis of the password strength among the given methods

Dependent variable: password strength						
Tukey HSD						
(*I*) method use	(*J*) method use	Mean difference (*I*–*J*)	Error	Std. sig.	95% Confidence interval	
					Lower bound	Upper bound
First method	Second method	2.32	1.700	0.525	− 2.11	6.74
	Third method	− 1.86	1.576	0.639	− 5.96	2.24
	None of them	14.63*	1.895	0.000	9.71	19.56
Second method	First method	− 2.32	1.700	0.525	− 6.74	2.11
	Third method	− 4.18*	1.589	0.047	− 8.31	− 0.04
	None of them	12.32*	1.906	0.000	7.36	17.28
Third method	First method	1.86	1.576	0.639	− 2.24	5.96
	Second method	4.18*	1.589	0.047	0.04	8.31
	None of them	16.50*	1.796	0.000	11.82	21.17
None of them	First method	− 14.63*	1.895	0.000	− 19.56	− 9.71
	Second method	− 12.32*	1.906	0.000	− 17.28	− 7.36
	Third method	− 16.50*	1.796	0.000	− 21.17	− 11.82

Based on observed means
The error term is mean square (error) = 49.817
* The mean difference is significant at the 0.05 level

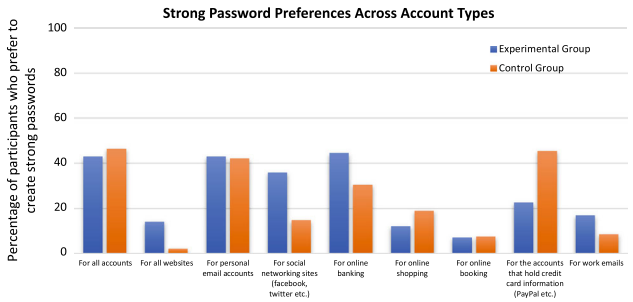

Fig. 17 Users' password preferences across different accounts

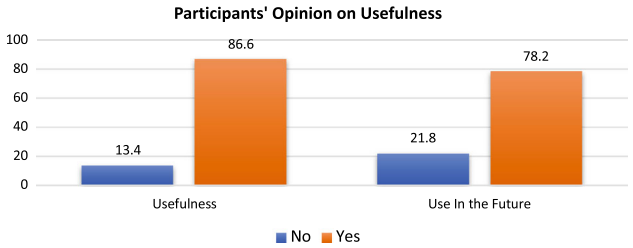

Fig. 18 Participants' opinion about the usability of the given methods

4.1.2 The results based on the survey responses

Password usage Among the participants who conducted the questionnaire, 80.3% of the experimental group and 66.3% of the control group reported that they did not experience any password security failure before. Moreover, only 39.4% of the experimental group and 31.6% of the control group reported that they do not write down their passwords anywhere, while 41.5% of the experimental group and 38.9% of the control group write their passwords somewhere safe and 19% of the experimental group and 29.5% of the control group prefer somewhere accessible.

Results also showed users' password preferences across different accounts vary. As seen in Fig. 17, most of the users tend to choose stronger passwords for their personal email accounts than the work emails.

User satisfaction Participants in the experimental group were asked about their experience of using the new methods presented to them. As illustrated in Fig. 18, most of the participants reported that they found the given methods useful to create a password for the current empirical study (87%) and they were likely to use given methods to create passwords in the future (79%).

Additionally, they rated their experience further in terms of fun and easiness. Most of the participants agreed that using the provided methods was easy (61%) and more than half of the participants evaluated the methods as fun to apply (54%). A larger body of participants (80%) reported that using the given methods to create secure passwords is worth the time spent on them. Moreover, most of the participants (85%)

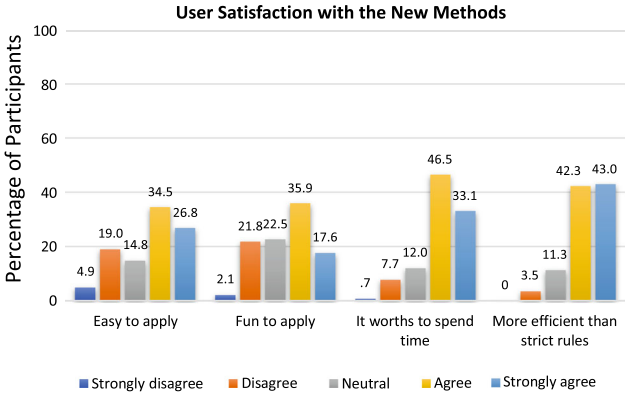

Fig. 19 User satisfaction with the new methods

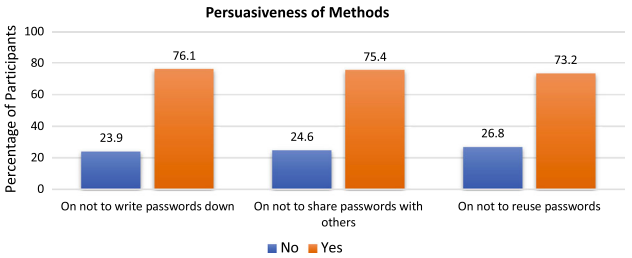

Fig. 20 The given methods' persuasiveness to abandon coping strategies

agreed that these methods were more efficient than the commonly used strict password creating rules, in terms of creating strong and memorable passwords (see Fig. 19).

Persuasiveness Finally, it seemed like the methods provided to the experimental group persuaded them not to use any of the coping strategies (see Fig. 20). Most of them reported they would not write down their passwords (76%), not share their passwords with other people (75%) and not reuse the passwords they created once (73%).

5 Discussion

The details of the web-based empirical study are presented in the previous sections. The aim of this study was to investigate whether users can create stronger and more memorable passwords if they are not enforced to comply with strict password policy rules. The study examined the efficiency of a persuasive text along with the three password composition methods on motivating users to create stronger and memorable passwords.

The results indicate that passwords created by users who receive password guidelines including a persuasive text and sample password creation methods are stronger compared to passwords created by those who are given the password guidelines including strict password composition rules. The

results also showed that the participants, who applied the given methods, remembered their passwords better than the ones who followed the usual password policy rules.

Thus, these findings suggest that it is a good idea for providing a message in the password guideline to persuade users to create their own password composition formula. When the message is supported with the example methods of password composition, it becomes possible to create strong passwords without burdening their memory to remember them. In this study, most of the participants in the experimental group spent time to read the information and applied the given methods to produce passwords, maybe just to help a research study by participating. However, in real life, users may not make an effort to read the information provided in the password guidelines unless they have to. Zakaria [64] suggested that one possible way to overcome this is to make reading and understanding the password guidelines compulsory before constructing a password. In particular, if the systems or applications require a high level of password security, this might be a reasonable solution. Another solution is visualisation of password attacks in password guideline to make users aware of the threats [67]. Ur et al. [56] stated that password advice should focus on promoting human algorithms for developing passwords, in addition to visualising threats in the password guideline. Some researchers claimed that users should be guided to save their limited mental capacity for passwords for high-value accounts [21,44,56].

The methods given in the password guidelines of the experimental group allowed users to create passwords which spontaneously meet the requirements of the password policy which was given to the control group as password guideline. In other words, most of the participants created passwords which are longer than 8 characters, including upper and lower cases, also numbers and special keyboard characters. However, the study showed that complying with these rules does not always guarantee creating strong passwords.

There are more attributes such as password length and not including meaningful words in passwords that affect the password strength. Therefore, two tools were used to measure password strength considering the frequency and variety of characters as well as the similarity to the dictionary words. Although, the results of the measurements performed with the *Password Meter* showed that the control group's passwords are less strong than the experimental group's, they were still in the "strong" category. However, the difference between the cracking times measured with the "How Secure is my Password" is huge. Most of the passwords of the control group categorised as strong seemed likely to be cracked in a day. The reason of this difference is that the second tool takes into consideration whether the password looks like a dictionary word besides measuring the character variability. The passwords including meaningful details such as dictionary words are likely to crack in short times even if they are composed of different characters. As a result, both measurements yielded results in support of password guideline including persuasive message and the example password creation methods.

Furthermore, questionnaire responses showed that the given methods are efficient to persuade users to abandon coping strategies to remember their passwords. Users in the experimental group seemed to be willing to comply with the password guideline and apply the given methods. However, a little more than half of the participants (54%) found the methods fun to apply. To increase the user satisfaction and lessen more the memory burden, images can be included in the password creation process. Users can also be supported by some feedback such as a traditional password meter or an emoji-based approach during the password selection. This might also make the password creation process fun for them [24].

There is a lack of ecological validity in generalising the findings to real life as the study is not conducted with a real application in use. The participants knew that the websites were created for a research study and they created passwords for helping the study. If users created passwords for a real website to perform their own tasks, quality of the passwords and compliance with the guidelines might have been different. On the other hand, the password guideline including several password creation tips provided to the experimental group is somewhat long to read. Therefore, users might not be willing to read it when they create a password in real life. This reduces the practicability of the proposed guideline. Another issue about the study is the unusual demographics of the experimental group. The participants were randomly assigned to the experimental and control groups. However, number of the participants who have high education level in the experimental group is more than the control group. This situation might have affected the password strength rates in the experimental group.

Moreover, the password creation methods provided to the users in the experimental group were not evaluated separately. Findings show that there was an apparent preference towards the third method so why it was more popular with the users could have also been investigated. Similarly, to point out the differences between the results of the participants from different countries could have been an interesting study. Unfortunately, in this study, the data and responses of all participants in the same group were evaluated together

6 Conclusion and further research

Although the use of passwords as an authentication method has been extensively studied in the past, there are no empirical studies that test the effectiveness of password creation

methods. Thus, the contribution of this research and its implications for both research and practice are significant.

As the results of the empirical studies indicated, the proposed password guideline improves the security and memorability of user-chosen passwords. Rather than obliging users to follow strict password policy rules, motivating and directing them to create strong and memorable passwords seems more efficient and usable way.

Thus far, the results generally have shown some promising findings; however, the practicability of this new password guideline may be an issue in real world, as stated above. The password guideline can be improved adding visual elements in the guideline to make reading the given information process interesting. It would probably be useful to attract users' attention and make the password creation process more enjoyable. Also, implementing the proposed password guideline into different kinds of applications which require different levels of security and conducting a further empirical study with different user groups involving more participants would be useful. Also, giving users some feedback during the password selection process such as meter-based ratings would motivate users to choose more secure passwords. Moreover, the literature on persuasion suggests that persuasion attempts are more likely to succeed if the persons are aware of the situation. Thus, adding some attributes to the password guideline informing users about possible attacks if they choose weak passwords might improve the compliance to password guideline.

References

1. Adams, A., Sasse, M.A.: Users are not the enemy. Commun. ACM **42**(12), 40–46 (1999). https://doi.org/10.1145/322796.322806
2. Barton, B.F., Barton, M.S.: User-friendly password methods for computer-mediated information systems. Comput. Secur. **3**(3), 186–195 (1984)
3. Bishop, M., Klein, D.V.: Improving system security via proactive password checking. Comput. Secur. **14**(3), 233–249 (1995)
4. Bonneau, J., Herley, C., van Oorschot, P.C., Stajano, F.: The quest to replace passwords: a framework for comparative evaluation of web authentication schemes. In: Security & Privacy (SP), IEEE Symposium, pp. 553–567 (2012)
5. Burnett, M., Kleiman, D. (eds.): Perfect Passwords. Syngress Publishing, Inc, Massachusetts (2006)
6. Burr, W., Dodson, D., Polk, W.: Electronic Authentication Guideline. Special Publication 800-63 Version 1.0.2. National Institute of Standards and Technology (NIST), Gaithersburg (2006)
7. Campbell, J., Ma, W., Kleeman, D.: Password composition policy: does enforcement lead to better password choices. In: Proceedings of the 17th Australian Conference on Information Systems, pp. 60–69 (2006)
8. Carstens, D.S., McCauley-Bell, P.R., Malone, L.C., DeMara, R.F.: Evaluation of the human impact of password authentication practices on information security. Inf. Sci. J. **7**, 67–85 (2004)
9. Carstens, D.S., Malone, L.C., McCauley-Bell, P.: Applying chunking theory in organizational password guidelines. J. Inf. Inf. Technol. Organ. **1**, 97–113 (2006)
10. Castelluccia, C., Duermuth, M., Perito, D.: Adaptive password-strength meters from Markov models. In: Network and Distributed System Security Symposium (NDSS, ISOC) (2012)
11. Chiasson, S., Biddle, R., van Oorschot, P.C.: A second look at the usability of click-based graphical password. In: Proceedings of the 3rd Symposium on Usable Privacy and Security, pp. 1–12. ACM (2007)
12. Chiasson, S., Forget, A., Stobert, E., van Oorschot, P., Biddle, R.: Multiple password interference in text and click-based graphical passwords. In: ACM Computer and Communications Security (CCS) (2009)
13. Cialdini, R.B.: Harnessing the science of persuasion. Harv. Bus. Rev. **79**(9), 72–81 (2001)
14. Clair, L.S., Johansen, L., Enck, W., Pirretti, M., Traynor, P., McDaniel, P., Jaeger, T.: Password exhaustion: predicting the end of password usefulness. In: International Conference on Information Systems Security, pp. 37–55. Springer (2006)
15. Collider, S.: How secure is my password? `Howsecureismypassword.net`. https://howsecureismypassword.net (2016). Accessed 14 Jan 2017
16. Colnago, J., Devlin, S., Oates, M., Swoopes, C., Bauer, L., Cranor, L., Christin, N.: "It's not actually that horrible": exploring adoption of two-factor authentication at a university. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, CHI '18, pp. 1–11. ACM, New York, NY, USA (2018)
17. Cowan, N.: The magical number 4 in short-term memory: a reconsideration of mental storage capacity. Behav. Brain Sci. **24**(1), 87–114 (2001)
18. Everitt, K.M., Bragin, T., Fogarty, J., Kohno, T.: A comprehensive study of frequency, interference, and training of multiple graphical passwords. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 889–898. ACM(2009)
19. Florencio, D., Herley, C.: A large-scale study of web password habits. In: Proceedings of the 16th International Conference on World Wide Web, WWW '07, pp. 657–666. ACM, New York, NY, USA (2007). https://doi.org/10.1145/1242572.1242661
20. Florencio, D., Herley, C.: Where do security policies come from? In: Proceedings of the Sixth Symposium on Usable Privacy and Security, SOUPS '10, pp. 1–14. ACM, New York, NY, USA (2010). https://doi.org/10.1145/1837110.1837124
21. Florencio, D., Herley, C., van Oorschot, P.C.: Password portfolios and the finite-effort user: sustainably managing large numbers of accounts. In: Proceedings USENIX Security, pp. 575–590 (2014)
22. Forget, A.: A world with many authentication scheme. Ph.D. thesis, Carleton University, Ottawa, Ontario (2012)
23. Furnell, S.: An assessment of website password practices. Comput. Secur. **26**(7), 445–451 (2007)
24. Furnell, S., Esmael, R., Yang, W., Li, N.: Enhancing security behaviour by supporting the user. Comput. Secur. **75**, 1–9 (2018)
25. Garrison, C.P.: An evaluation of passwords. CPA J. **78**(5), 70 (2008)
26. Goldberg, J., Hagman, J., Sazawal, V.: Doodling our way to better authentication. In: CHI Extended Abstracts on Human Factors in Computing Systems, pp. 868–869. ACM (2002)
27. Grassi, P.A., Garcia, M.E., Fenton, J.L.: Digital Identity Guidelines. Special Publication (NIST SP) 800-63-3. NIST, Gaithersburg (2017)
28. Grawmeyer, B., Johnson, H.: Using multiple password: a week to a view. Interact. Comput. **23**(3), 256–267 (2011)

29. Habib, H., Colnago, J., Melicher, W., Ur, B., Segreti, S.M., Bauer, L., Christin, N., Cranor, L.F.: Password creation in the presence of blacklists. In: Workshop on Usable Security, USEC '17, Internet Society (2017)
30. Herley, C., van Oorschot, P., Patrick, A.: Passwords: If we're so smart, why are we still using them? In: Dingledine, R., Golle, P. (eds.) Financial Cryptography and Data Security, FC 2009. Lecture Notes in Computer Science, vol. 5628, pp. 230–237. Springer, Berlin (2009)
31. Hub, M., Čapek, J., Myšková, R.: Relationship between security and usability-authentication case study. Int. J. Comput. Commun. **5**(1), 1–9 (2011)
32. Inglesant, P.G., Sasse, M.A.: The true cost of unusable password policies: Password use in the wild. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '10, pp. 383–392. ACM, New York, NY, USA (2010). https://doi.org/10.1145/1753326.1753384
33. Jeyaraman, S., Topkara, U.: Have the cake and eat it too - infusing usability into text-password based authentication systems. In: Proceedings of the 21st Annual Computer Security Applications Conference, ACSAC '05, pp. 473–482. IEEE Computer Society, Washington, DC, USA (2005). https://doi.org/10.1109/CSAC.2005.28
34. Keith, M., Shao, B., Steinbart, P.J.: The usability of passphrases for authentication: an empirical field study. Int. J. Hum. Comput. Stud. **65**(1), 17–28 (2007)
35. Kelley, P.G., Komanduri, S., Mazurek, M.L., Shay, R., Vidas, T., Bauer, L., Christin, N., Cranor, L.F., Lopez, J.: Guess again (and again and again): measuring password strength by simulating password-cracking algorithms. In: Proceedings of the 2012 IEEE Symposium on Security and Privacy, pp. 523–537. IEEE (2012)
36. Klein, D.V.: Foiling the cracker: a survey of, and improvements to, password security. In: Proceedings of the 2nd USENIX Security Workshop, pp. 5–14 (1990)
37. Knieriem, B., Zhang, X., Levine, P., Breitinger, F., Baggili, I.: An overview of the usage of default passwords. In: Matoušek, P., Schmiedecker, M. (eds.) Digital Forensics and Cyber Crime. ICDF2C 2017, Springer, Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol. 216, pp. 195–203 (2018)
38. Komanduri, S., Shay, R., Kelly, P.G., Mazurek, M.L., Bauer, L., Christin, N., Egelman, S.: Of passwords and people: Measuring the effect of password-composition policies. In: Proceedings of the Human Factors and Computing Systems, pp. 2595–2604. ACM (2011)
39. Kotadia, M.: Gates predicts death of the password. https://www.cnet.com/news/gates-predicts-death-of-the-password/ (2014). Accessed 10 May 2014
40. Manning, C.D., Schutze, H.: Foundations of Statistical Natural Language Processing. The MIT press, Cambridge (1999)
41. Miller, G.A.: The magical number seven, plus or minus two: some limits on our capacity for processing information. Psychol. Rev. **63**(2), 81 (1956)
42. Moncur, W., Leplâtre, G.: Pictures at the atm: exploring the usability of multiple graphical passwords. In: Proceedings of the SIGCHI conference on Human factors in computing system, pp. 887–894 (2007)
43. Murray, H., Malone, D.: Evaluating password advice. In: 28th Irish Signals and Systems Conference (ISSC), pp. 1–6 (2017)
44. Nithyanand, R., Johnson, R.: The password allocation problem: strategies for reusing passwords effectively. In: Proceedings of the 12th ACM Workshop on Workshop on Privacy in the Electronic Society, WPES '13, pp. 255–260. ACM, New York, NY, USA (2013). https://doi.org/10.1145/2517840.2517870
45. Passwordmetercom: password strength checker. http://www.passwordmeter.com (2017). Accessed 5 Jan 2017
46. Perloff, R.M.: The Dynamics of Persuasion. Taylor and Francis, London (2016)
47. Proctor, R.W., Lien, M.C., Vu, K.P.L., Schultz, E.E., Salvendy, G.: Improving computer security for authentication of users: influence of proactive password restrictions. Behav. Res. Methods **34**(2), 163–169 (2002)
48. Roig, J.: Do smarter people have better passwords? (2018). arXiv:1805.02931
49. Shannon, C.E.: Prediction and entropy of printed english. Bell Syst. Tech. J. **30**(1), 50–64 (1951). https://doi.org/10.1002/j.1538-7305.1951.tb01366.x
50. Shay, R., Komanduri, S., Kelley, P.G., Leon, P.G., Mazurek, M.L., Bauer, L., Christin, N., Cranor, L.F.: Encountering stronger password requirements: User attitudes and behaviors. In: Proceedings of the Sixth Symposium on Usable Privacy and Security, SOUPS '10, pp. 1–20. ACM, New York, NY, USA (2010). https://doi.org/10.1145/1837110.1837113
51. Shay, R., Kelley, P.G., Komanduri, S., Mazurek, M.L., Ur, B., Vidas, T., Bauer, L., Christin, N., Cranor, L.F.: Correct horse battery staple: Exploring the usability of system-assigned passphrases. In: Proceedings of the Eighth Symposium on Usable Privacy and Security, SOUPS '12, pp. 1–20. ACM, New York, NY, USA (2012). https://doi.org/10.1145/2335356.2335366
52. Summers, W., Bosworth, E.: Password policy: the good, the bad, and the ugly. In: Proceedings of the Winter International Symposium on Information and Communication Technologies, pp. 1–6. ACM (2004)
53. Thorpe J, MacRae, B., Salehi-Abari, A.: Usability and security evaluation of geopass: A geographic location-password scheme. In: Proceedings of the Ninth Symposium on Usable Privacy and Security, SOUPS '13, pp. 1–14. ACM, New York, NY, USA (2013). https://doi.org/10.1145/2501604.2501618
54. Topkara, U., Topkara, M., Atallah, M.J.: Passwords for everyone: secure mnemonic-based accessible authentication. In: USENIX Annual Technical Conference, pp. 369–374 (2007)
55. Ur, B., Kelley, P.G., Komanduri, S., Lee, J., Maass, M., Mazurek, M.L., Passaro, T., Shay, R., Vidas, T., Bauer, L., Christin, N., Cranor, L.F.: How does your password measure up? The effect of strength meters on password creation. In: Presented as Part of the 21st USENIX Security Symposium (USENIX Security 12), USENIX, Bellevue, WA, pp. 65–80 (2012)
56. Ur, B., Noma, F., Bees, J., Segreti, S.M., Shay, R., Bauer, L., Christin, N., Cranor, L.F.: "i added '!' at the end to make it secure": Observing password creation in the lab. In: Eleventh Symposium On Usable Privacy and Security (SOUPS 2015), USENIX Association, Ottawa, pp. 123–140 (2015)
57. Vu, K.P.L., Proctor, R.W., Bhargav-Spantzel, A., Tai, B.L.B., Cook, J., Eugene Schultz, E.: Improving password security and memorability to protect personal and organizational information. Int. J. Hum. Comput. Stud. **65**(8), 744–757 (2007)
58. Weir, M., Aggarwal, S., Collins, M., Stern, H.: Testing metrics for password creation policies by attacking large sets of revealed passwords. In: Proceedings of the 17th ACM Conference on Computer and Communications Security, CCS '10, pp. 162–175. ACM, New York, NY, USA (2010)
59. Wiedenbeck, S., Waters, J., Birget, J.C., Brodskiy, A., Memon, N.: Authentication using graphical passwords: effects of tolerance and image choice. In: Proceedings of the 2005 Symposium on Usable Privacy and Security, SOUPS '05, pp. 1–12. ACM, New York, NY, USA (2005)
60. Wood, D., Bruner, J.S., Ross, G.: The role of tutoring in problem solving. J. Child Psychol. Psychiatry **17**(2), 89–100 (1976)
61. Wright, N., Patrick, A.S., Biddle, R.: Do you see your password?: Applying recognition to textual passwords. In: Proceedings of the Eighth Symposium on Usable Privacy and Security, SOUPS '12, pp. 8:1–8:14. ACM, New York, NY, USA (2012)

62. Yan, J., Blackwell, A., Anderson, R., Grant, A.: Password memorability and security: empirical results. IEEE Priv. Secur. **2**(5), 25–31 (2004)
63. Yan, J.J.: A note on proactive password checking. In: Proceedings of the 2001 Workshop on New Security Paradigms, NSPW '01, pp. 127–135. ACM, New York, NY, USA (2001)
64. Zakaria, N.H.B.: Exploring human factors issues & possible countermeasures in password authentication. Ph.D. thesis, Newcastle University (2013)
65. Zhang, L., McDowell, C.W.: Am I really at risk? Determinants of online users' intention to use strong passwords. J. Internet Commer. **8**, 180–197 (2009)
66. Zhang, Y., Monrose, F., Reiter, M.K.: The security of modern password expiration: an algorithmic framework and empirical analysis. In: Proceedings of the 17th ACM Conference on Computer and Communications Security, pp. 176–186 (2010)
67. Zhang-Kennedy, L., Chiasson, S., Biddle, R.: Password advice shouldn't be boring: visualizing password guessing attacks. In: 2013 APWG eCrime Researchers Summit, pp. 1–11 (2013)
68. Zviran, M., Haga, W.J.: Comparison of password techniques for multilevel authentication mechanisms. Comput. J. **36**(3), 227–237 (1993)